Author's response to reviews

**Title:** Can electronic search engines optimize screening of search results in systematic reviews: an empirical study

**Authors:**

- Margaret Sampson (msampson@uottawa.ca)
- Nicholas J Barrowman (nbarrowman@cheo.on.ca)
- David Moher (dmoher@cheo.on.ca)
- Tammy J Clifford (tammyc@cchota.ca)
- Robert W Platt (robert.platt@mcgill.ca)
- Andra Morrison (andram@ccohta.ca)
- Terry Klassen (terry.klassen@ualberta.ca)
- Li Zhang (li.zhang@usask.ca)

**Version:** 2 **Date:** 22 December 2005

**Author's response to reviews:** see over
Reply to reviewers
Thank you for your helpful review of this manuscript. We have addressed all suggested revisions and the manuscript is improved as a result. Details of revisions are below.

Sincerely,
Margaret Sampson

Major Compulsory Revisions
The Objective listed just before the Methods section is not at all clear or concise. The objective in the abstract is much clearer. Please modify the Objective in the body of the work.

Thank you. We have replaced the objective in the body of the work with that stated in the abstract.

Minor Essential Revisions
All commercial products and services need to have their trademark/copyright information included. Please add all especially ones associated with MEDLINE and Ovid. Here is the information relating to MEDLINE.
Please give a description of and the contact information on Ultraseek® and maybe a web address. Also do the same for SRS. I am not sure all your readers will know these 2 products.

Thank you for this information. We have added the ® to the first instance of MEDLINE, Ultraseek, DIALOG and Reference Manager (Reference Manager had been TM) we added TM to Ovid, SRS. There seems to be some variation between style manuals regarding proper use of these symbols. Some would argue that such usage is not necessary, regardless of the wishes of these holding the trademarks (http://www.eeicom.com/eye/careful.html). We could not find ICMJE instructions on this point. We have followed the IEEE Standards Style Manual (http://standards.ieee.org/guides/style/)

After the first mention of SRS we have added this sentence:
SRS is a web-based platform for conducting systematic reviews produced by TrialStat (http://www.trialstat.com/).
Before the first mention of Ultraseek in the Search Engine Configuration section within Methods we have added this sentence: “Produced by Verity, Ultraseek® was originally a successful web search engine, but is now focused on helping businesses manage their digital information (http://www.verity.com/products/search/ultraseek/.)”

While we were at it, we changed several instance of lowercase html to uppercase so that uppercase is used whenever HTML is free-standing (not part of code or a file name). Usage had been inconsistent in the previous version.
Discretionary revisions:
Thank you for the many helpful suggestions to improve the clarity of the manuscript.

The full reports are generally obtained for the remaining items, and reviewers make a final determination of eligibility after examining the complete full-text document.

Changed to “In the second stage, reviewers usually obtain the full articles associated with remaining records and then decide eligibility based on the complete report, rather than on the more limited information available at the first stage of screening.”

First paragraph.....primary studies that were either aN rct or

Changed to “have been a reviews or RCTs or quasiRCTs.”

…the phrase (hsss or highly sensitive search).tw......Please clarify. I think I know what you are saying but I am not sure.

Changed to “using the search string (hsss or highly sensitive search).tw. to identify potential studies.”

…data about the eligible cases…. First, and last time you use “cases”.
Very true – we have changed this to “the eligible reviews”

I had trouble sorting out a “generic” review and one of the 9 “case-reviews”. Please consider using some sort of terminology to sort out in the readers mind when you talk about generic reviews and your case-reviews. Collection is a term you use that might be useful here too.

We have changed the heading “Results of the full search replication” to “Results of the full search replication of 9 reviews”

The indexing status of each study was recorded
We changed this to “The indexing status of each study was recorded as indexed or not indexed”

Search engine configuration
I think that Ultraseek processed all citations up to 6000 and then displayed what it thought to be the 500 most highly relevant citations. Please clarify this process and include in this section.

You are correct. The answer was buried the legend to a figure. To be clearer, we have added the following description of the process to this section:
“The search engine indexed all records in the collection. When a search was run against the collection, the number of items with relevance greater than zero was returned, along with a listing of the first 500 relevant items, sorted by relevance.”
In classic information retrieval studies, the greatest weight is often given to title words and not indexing terms. Could you please cite a reference that supports your choice of more weighting to the index terms?

We have provided a bit more explanation and a reference:

“As we were dealing with indexed records and indexers has the benefit of access to the full text of the document, we anticipated that relevance ranking could be optimized by assigning greatest weight to terms appearing in indexing fields, intermediate weight to terms appearing in the title field, and lowest weight to terms appearing in the abstract field, following Hutchinson. [8]”


Search terms
You have Cochrane/MEDLINE search terms and you have Ultraseek search terms. Could you somehow indicate which ones you are talking about in the text of the article?

…”reasonable separation”—can you use a different term here or more explanation?

Yes, we would be happy to: “Thus we decided to standardized our Ultraseek searches at 7 terms, the minimum number that seemed to reduce ties to a workable number.

“random*”—can you add that the star means truncation (if that is what is meant)
Thank you for pointing out this ambiguity. We have added this second sentence that we hope clarifies this:
A final eighth term, “random*” was included in each search. The asterisk is the truncation symbol used with Ultraseek.

Analysis
…we also examined the recall of included studies into the top 10 ranks
…we also examined the recall of included studies by evaluating only those studies that Ultraseek identified as being the 10 most relevant studies (top 10 ranks) for each case-review… (or something like that?)
How about precision? Precision being number of relevant articles retrieved divided by the number recalled – 10 in this case. We have changed the text to read: we also examined the precision of the top 10 rankings. [12]

Results of the full search replication
For three case-reviews where the ranking performed poorly, we attempted another Ultraseek(?) search...
Thank you –changed to “For three reviews where the Ultraseek ranking performed poorly,…”

Similarly, no obvious pattern between the included and excluded studies could be discerned. This sentence is not clear to me nor is it clear how it related to the previous sentence.
Indeed. We hope this is clearer: “Further, no obvious pattern between the included and excluded studies could be discerned through visual inspection (see for example, Figure 2).”

Systematic reviewers will need to become even more innovative if reviews are to maintain their current status. (Why? Not sure I follow the logic here.)

Thank you for catching this. In version 1 of this manuscript we failed to mention that timeliness was a key factor for policy makers. We have rectified this by amending the sentence to read “Systematic reviewers will need to become ever more innovative if reviews are to maintain the timeliness necessary for decision makers” and we have supported this with references.

You used a term weighting approach to relevance ranking. Your discussion could include a sentence that states that other methods of ranking (e.g., vector space methods, probabilistic methods, machine learning algorithms, and natural language processing procedures) have been used in information retrieval studies and they each could be used as possible alternatives to identifying studies for systematic reviews. James Cimino (2000s) and Bill Hersh (early to mid 1999s) have used natural language processing methods in medical settings to rank outputs—still not perfect.

One disadvantage of using proprietary software is that we do not know the precise mechanism used for indexing and ranking (it may be that vector space methods are used). We can comment on the problems with using key words for retrieval and have added these sentences to our discussion of forming keyword queries.

Further, using a limited number of terms as keywords for retrieval prevents the formation of elaborate queries,[28] and we experienced difficulty representing some of the topics of the reviews. Other approaches, such as natural language queries, have been explored as alternatives to Boolean retrieval of medical literature.[29,30]

Reviewer: Arminee Kazanjian
Discretionary Revisions (which the author can choose to ignore)
It would be interesting to raise and discuss, if not actually measure, the impact of one person identifying subject terms for the Ultraseek® search on relevance scores. This especially as the authors note the relationship between number of terms entered and tied relevance scores.

We agree that this might be problematic. Had we achieved more compelling results we would have wanted to measure the degree to which different people to get the search engine to perform.

We have moved the statement in the discussion addressing this to the end of the paragraph and changed it from "A limitation was the dependence on term selection and even the ordering of terms." To "A limitation was the sensitivity to
term selection and even the ordering of terms, thus performance could depend on operator characteristics."